scripted

Volume 11, Issue 3, December 2014

SHARE AND SHARE ALIKE?

AN EXAMINATION OF TRUST, ANONYMISATION AND DATA SHARING WITH PARTICULAR REFERENCE TO AN EXPLORATORY RESEARCH PROJECT INVESTIGATING ATTITUDES TO SHARING PERSONAL DATA WITH THE PUBLIC SECTOR

*Marion Oswald**

Abstract

This article asks whether the necessity of many public services results in a readiness of individuals to share personal data, and thus sacrifice a certain level of privacy, in connection with their provision. It will explore the value of privacy in the context of the on-going debates around personal data sharing, with particular focus on the public sector in England, using the UK government's care.data project as an example. The impact on trust relations between the government, the National Health Service (NHS) and the citizen will be considered. The importance of anonymisation of personal data as a method of minimising privacy risks and increasing trust will be discussed. Using the results of the author's exploratory empirical study into attitudes to sharing personal data with the public sector, the article will suggest that the benefits-versus-costs privacy problem is particularly significant in relation to data sharing projects in the public sector. The lack of definitive answers in relation to the risk of re-identification contributes to the problem. Finally, the article will suggest that future work may wish to investigate how trust in, and acceptance of, data sharing initiatives could be improved by a bottom-up institution-led approach.

DOI: 10.2966/scrip.110314.245

* Marion Oswald, Senior Fellow, Head of the Centre for Information Rights, University of Winchester marion.oswald@winchester.ac.uk . Research funded by a research grant from the British and Irish Law Education and Technology Association (BILETA). I am grateful to Matthew Jordan, Samantha Borek and Emma Vinson for their research assistance.

1. Introduction

The collection of personal data and disclosure of that data by one party to another is often the lifeblood of public services, with those services assuming "the availability of a substantial quantity of personal data and hence, a readiness by an individual to supply it."[1] Can it be concluded, however, that the necessity of many of these services results in such a readiness in individuals to share personal data, and thus sacrifice a certain level of privacy, in connection with their provision? The use of personal data by the public sector is no longer confined to personalised, localised services; it can be used to create open datasets "to harvest unused knowledge that otherwise goes to waste, which can be used to empower citizens, to improve public services, and to benefit the economy and society as a whole."[2] And personal data can be amalgamated into Big Data which "arrives with big promises:"[3] among other things, to predict 'trouble spots and troublemakers,' combat fraud by detecting anomalous behaviour patterns[4] and to drive up the quality of care in the healthcare system.[5] As such, justification for broader use of personal data within the public sector may, controversially, be moving away from traditional models of data collection based on individual consent to ones based on an overriding public interest or necessity, on the basis that "when much of data's value is in secondary uses that may have been unimagined when the data was collected,' a 'formulaic system of 'notice and consent'" is no longer suitable.[6] Such an approach may be combined with anonymisation of the personal data, resulting in the claim that data *sharing* (as the disclosure and acquisition of data has become known) therefore has no privacy impact.

But legitimate privacy concerns remain. Collection and dissemination of personal data by the public sector raises the spectre of potential harms to the individual: "architectural problems" as Solove called them, upsetting the balance of social and institutional power; distortion,

---

[1] R. Wacks, *Privacy: A very short introduction* (1st edn, Oxford University Press, Oxford 2010), at 110.

[2] House of Commons Public Administration Committee, *Statistics and Open Data: Harvesting unused knowledge, empowering citizens and improving public services*, *Tenth Report of Session* HC 564 (London: The Stationary Office, 2014), at 3 available at http://www.publications.parliament.uk/pa/cm201314/cmselect/cmpubadm/564/56402.htm (last accessed 24 March 2014).

[3] W. Davies, *Empirical Limits*,(2013) 4 RSA Journal 36-39, at 36.

[4] M. Ward, *Crime fighting with big data weapons* BBC (18 March 2014) http://www.bbc.co.uk/news/business-26520013 (last accessed 24 March 2014).

[5] J. Lewis, *Putting care.data into context* StatsLife (17 February 2014) available at http://www.statslife.org.uk/opinion/1246-putting-care-data-into-context (last accessed 24 March 2014).

[6] V. Mayer-Schonberger and K. Cukier, *Big Data*, (John Murray, 2013) at 173.

where data fails to mirror the entire individual and his or her personal circumstances; and exclusion of individuals from decisions about how their information is used.[7] Individuals may choose to pay the "informational price"[8] of disclosing personal data to social media and search engines, because the immediate benefits outweigh the potential harms, the social cost to the individual of opting-out is too high and the "notice and consent" model gives the individual, at the very least, a semblance of control. The dilemma for the public sector is how to demonstrate that long-term societal aims outweigh potential harms and to counter the view that "nobody will be as comfortable with the idea of government as data analyst as they are with the idea of Netflix or, to a lesser extent, Facebook tracking our behaviour *en masse*."[9]

2. Structure

The structure of this article is as follows. First, the article will explore the value of privacy in the context of the on-going debates around data sharing, with particular focus on the public sector in England. On the one hand, there are numerous reports in which the public sector has been criticised for failing to share personal data effectively.[10] On the other, there are real concerns around the aggregation of personal data, data mining and profiling and data intelligence techniques, giving rise for example to the current debate around the 'care.data' project in England[11] and the impact on trust relations between the government, the National Health Service (NHS) and the citizen.

Secondly, the anonymisation of personal data as a method of minimising privacy risks and increasing trust will be reviewed. Questions as to whether anonymity can ever be guaranteed have been important to the care.data project and this section will consider how the risk-based nature of personal data anonymisation may contribute to privacy concerns.

---

[7] D.J. Solove, "A Taxonomy of Privacy" (2006) 154 (3) *University of Pennsylvania Law Review* 477-560

[8] H. Nissenbaum, *A Contextual Approach to Privacy Online*, (2011) 140(4) *Journal of the American Academy of Arts & Sciences* 32-48, at 35.

[9] W. Davies, *Empirical Limits*, (2013) 4 *RSA Journal* 36-39, at 36.

[10] House of Commons, *The Bichard Inquiry Report* , HC653 (June 2004); M. Flynn *South Gloucestershire Safeguarding Adults Board, Winterbourne View Hospital, A Serious Case Review* (July 2012); Her Majesty's Inspectorate of Constabulary, *Mistakes were made* (2013); Coventry Safeguarding Children Board, *Serious Case Review Re Daniel Pelka* (September 2013); Birmingham Safeguarding Children Board, *Serious Case Review in respect of the death of Keanu Williams* (September 2013); National Crime Agency, *CEOP Thematic Assessment, The Foundations of Abuse: A thematic assessment of the risk of child sexual abuse by adults in institutions* (October 2013); S. Berelowitz et al, *"If only someone had listened" Office of the Children's Commissioner's Inquiry into Child Sexual Exploitation in Gangs and Groups, Final Report* (November 2013)

[11] See http://www.hscic.gov.uk/patientconf .

Next, a recent exploratory empirical research project undertaken by the author will be discussed. The aim of the study was to bring empirical facts to questions of public confidence in data sharing practice by investigating people's attitudes to sharing personal data with the public sector, in particular, attitudes towards local councils, central government and the NHS and focussing on three categories of personal data: locational data, minor personal ailments and detailed medical history. Anonymisation of personal data was a particular focus of the study. It was hypothesised that participants would have significantly different comfort levels with providing their personal data to the selected public sector organisations. It was also hypothesised that there would be a difference in participants' comfort levels between personal data being collected, stored and used, and data being shared, and that anonymisation would increase comfort levels with data sharing. The study also explored comfort levels in relation to the public sector using or sharing data for different reasons.

The concluding section considers how the lack of measurable and definite principles, particularly as regards the anonymisation of personal data, may contribute to a lack of trust and misconceptions, and how the research results might support the application of localism principles to future data sharing initiatives. Anyone involved in such initiatives may wish to consider the suggested avenues for further investigation.

3. Personal data sharing and privacy: the public sector dilemma

O'Hara neatly summed up the dilemma facing many public bodies: "The point about privacy is that it raises hard cases; people want privacy for perfectly good reasons, and others want information for equally good reasons."[12] Collection of personal data, and transfer of such data from one public sector body to another, within a public sector body, or to the private sector can be of fundamental importance to the successful delivery of public services, the identification of risk and the discharge of government responsibilities. Yet, as identified by the Law Commission, "a low public acceptance of data sharing and a low level of trust in the way it is undertaken by public services, along with negative media coverage" may create hindrances to sharing.[13] Commenting on issues relating to medical research, Love and Sullivan highlighted the climate of uncertainty stemming from the interpretative nature of the Data Protection Act and Human Rights Act: "custodians of health information might feel that

[12]K. O'Hara and N. Shadbolt, *The Spy in the Coffee Machine* (One World Publications, 2008) at 23.

[13] Law Commission Consultation Paper No 214, *Data Sharing Between Public Bodies* (2013) at 6.

the only way to protect themselves from future allegations of impropriety with data is to take the most conservative option when considering requests from researchers."[14]

The privacy debate around personal data sharing is often couched in terms of achieving a balance between privacy and the other value or aim, implying some form of trade-off between the two. It may also suggest an all-or-nothing situation, i.e. either you have privacy or you achieve the other aim. "The privacy paradigm encourages the view that individual privacy and social values such as sociability, internal security, social welfare or government efficacy are necessarily antithetical."[15] Participants in the debate tend to bring to any assessment of personal data sharing their own assumptions as to whether privacy is good or bad (or whether the competing aim is good or bad) rather than beginning with a neutral conception of privacy. Nissenbaum commented that starting with such a neutral conception "leaves open the possibility that in certain circumstances less privacy might be better than more and that reductions in privacy need not constitute violations or intrusions or incursions."[16] Raab agreed: "The paradigmatic emphasis on procedural due process and on an individualist construct of the value of privacy militates against raising distributional issues of privacy protection."[17] In the debate about data sharing with, and within, the public sector, it would be valuable, as Raab suggested, to have space to debate whether an "uneven distribution of data protection" may be justifiable.[18]

3.1 Failures of data sharing

Healthcare and safeguarding of the vulnerable are both fields in which less privacy (within the boundaries of the relevant public services) may in certain cases be better than more. In these fields, data sharing can be vital for monitoring quality, detecting abuses and for research purposes. It might therefore be expected that personal data sharing arrangements in these fields would be relatively uncontroversial, both in terms of the willingness of public bodies to share personal data and in terms of individuals' acceptance of such arrangements.

---

[14] T. Love and F. Sullivan, "Confidentiality, clinical governance and research in the community" (2004) 12(1) *Informatics in Primary Care* 1-2 at 1.

[15] C.D. Raab, "The future of privacy protection, in Trust and Crime" in R. Mansell and B.S. Collins, *Information Societies* (Edward Elgar Publishing, 2005) at 288.

[16] H. Nissenbaum, *Privacy in Context, Technology, Policy and the Integrity of Social Life*, (Stanford University Press, 2010) at 68.

[17] C.D. Raab, "The future of privacy protection, in Trust and Crime" in R. Mansell and B.S. Collins, *Information Societies* (Edward Elgar Publishing, 2005) at 289.

[18] *Ibid*, at 290.

Review after review, however, continues to criticise the lack of robustness in data sharing arrangements between public bodies. For instance, the Serious Case Review into the abuse committed against vulnerable adults at the Winterbourne View Care Home noted that:

- South Gloucestershire Council Adult Safeguarding received 27 allegations of staff to patient assaults, 10 allegations of patient on patient assaults, and 3 family related alerts;[19]
- Avon and Somerset Constabulary recorded 9 carer on patient incidents, 5 patient on patient incidents, 3 patient on carer incidents and 12 other incidents;[20]
- Between January 2008 and May 2011, patients attended Accident & Emergency on 76 occasions, yet there were no safeguarding alerts from A&E;[21]
- Castlebeck Ltd. (the company which owned the home) recorded 379 physical interventions during 2010 and 129 for the 1st quarter of 2011;[22]
- Information submitted to the Health and Safety Executive by Castlebeck was neither known to the Care Quality Commission nor to the relevant adult safeguarding team.[23]

The report concluded that drawing together all this information, together with complaints from patients and parents, would have identified the risks to which patients at Winterbourne View were subject: "Given that many patients were isolated and disconnected from sustaining relationships, the case for aggregating such information sources is compelling."[24] Recommendations included the introduction of a mechanism for aggregating pertinent safeguarding information for NHS patients with learning disabilities and autism,[25] and the exploration of how A&E could detect instances of re-attendance from the same location as well as by an individual.[26]

---

[19] M. Flynn, *South Gloucestershire Safeguarding Adults Board, Winterbourne View Hospital, A Serious Case Review* (July 2012), at 131 available at http://hosted.southglos.gov.uk/wv/report.pdf (last accessed 24 March 2014).

[20] *Ibid*, at 110.

[21] *Ibid*, at 134.

[22] *Ibid*, at 131.

[23] *Ibid*, at 132.

[24] *Ibid*, at 133.

[25] *Ibid*, at 132.

[26] *Ibid*, at 135.

3.2 Intelligence and privacy

Such aggregated and analysed information is *intelligence*, that is, information that may increase our understanding of a certain issue or problem. Intelligence held by public authorities, particularly in the context of health and social care and the prevention of crime, is often subjective as well as partial and unverifiable, and its analysis may be speculative with no guarantee of a useful outcome resulting. The potential unreliability of intelligence makes the decision as to whether or not to share a difficult one. Yet intelligence can be fundamental to the accurate identification of risk and, as mentioned above, Big Data promises much in this area. The Francis Review into the failures of care at Mid-Staffordshire NHS Trust commented that General Practitioners "need to have internal systems enabling them to be aware of patterns of concern, so that they do not merely treat each case on its individual merits."[27] The Daniel Pelka Serious Case Review commented that "Instances of concern tended to be viewed in isolation with a lack of attention to the patterns developing."[28]

The police, the intelligence agencies and those involved in fraud prevention will be more familiar than other public authorities with the need to handle and share intelligence *intelligently*, and the potential adverse consequences if this is not done. The Bichard Report (which investigated police intelligence and data sharing failures in connection with the murderer Ian Huntley) commented that "two incidents taking place some time apart, but involving the same alleged offender, might be graded at the unreliable end of the scale when assessed separately. However, the connection between the same alleged offender and each incident might be an important intelligence item."[29] Or as Sherlock Holmes put it, "there is nothing so important as trifles."[30]

In the post-Snowden era, however, the term *intelligence* is in danger of becoming a discredited word. The generation of intelligence may involve the use of what are sometimes referred to as 'bulk' or 'Big' datasets in order to enrich existing or new information by, for instance, comparing one dataset to another or performing a cluster analysis (and the datasets required for such analysis will often have been shared between one public sector body and another, or by the private sector with the public sector). The proportionality of this type of activity can be called into question because of (among other things) the potentially large number of 'innocent' individuals contained within the bulk datasets, the volume and extent of the data collected (whether or not subsequently analysed), the potential vulnerability of the

---

[27] Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry Executive Summary, (London: The Stationery Office, 2013) HC 947, at 48.

[28] R. Lock, *Coventry Safeguarding Children Board, Serious Case Review Re Daniel Pelka Overview Report* (September 2013), at 69.

[29] House of Commons *The Bichard Inquiry Report*, HC653 (June 2004) at 135.

[30] Arthur Conan Doyle, *The Man with the Twisted Lip* (1891).

dataset to loss or malicious attack, the perceived speculative nature of the analysis being undertaken and because of the extensive profile of an individual that may be created. "People expect certain limits on what is known about them and on what others will find out. Aggregation upsets these expectations, because it involves the combination of data in new, potentially unanticipated ways to reveal facts about a person that are not readily known."[31]

Snowden's revelations concerning the bulk collection of internet meta-data by security agencies illustrate the consequences of upsetting those expectations, or as Nissenbaum put it, breaching "informational norms."[32] The immediate damage to individuals may be difficult to quantify, yet the concerns are real. Bernal commented that "'new' surveillance is both qualitatively and quantitatively different from what might be labelled 'traditional' surveillance or interception of communications. This means that the old debates, the old balances, need to be recast."[33] A linked concern relates to the extent of data that was being collected; was it more than could be used effectively?

This is a concern that not only relates to surveillance by the State but also to bulk data collection and analysis by any public body. According to Nissenbaum, the hope placed in data mining and profiling "has catapulted information – raw and processed – into a dynamic, starring role in social decision making" creating a "virtually unquenchable thirst" for information.[34] Does the pursuit of information satisfy a genuine need or is it "some sort of displacement activity" used to delay making difficult decisions, as asserted by Kennedy in his review into the breast care scandal at Solihull hospital? [35] Are we in danger of falling victim to "a dictatorship of data, whereby we fetishise the information, the output of our analyses, and end up misusing it"[36]?

Of course, what some may see as a speculative activity may be viewed by others as an important intelligence gathering activity; as those involved in healthcare research and crime

---

[31] D.J. Solove, *A Taxonomy of Privacy*, University of Pennsylvania Law Review, (2006) Vol. 154 No. 3 477-560, at 507.

[32] See note 16 above, at 227.

[33] P. Bernal, Submission to the Intelligence and Security Committee, (2014) available at http://paulbernal.wordpress.com/2014/02/06/communications-surveillance-a-miscast-debate/ (last accessed 11 March 2014)

[34] See note 16 above, at 44.

[35] I. Kennedy, *Review of the Response of Heart of England NHS Foundation Trust to Concerns about Mr Ian Paterson's Surgical Practice; Lessons to be Learned; and Recommendations*, (2013) http://www.heartofengland.nhs.uk/wp-content/uploads/Kennedy-Report-Final.pdf at 117.

[36] V. Mayer-Schonberger and K. Cukier, *Big Data* (John Murray, 2013) at 151.

detection know well, it may not be possible to identify suspicious or valuable patterns if the data is not comparable to the whole. In addition, it is sometimes as important to eliminate links as it is to identify them. Davies has said that "One of the dangers lurking in the promise of big data is that we will be governed and managed with the assertion that everything is empirically valid, but without being able to know what premises or principles are at work in how data has been scraped or trials been conducted."[37] Transparency that results in increased knowledge and understanding of the reasons for a data gathering and sharing exercise, the principles behind how the data will be amalgamated and analysed, and any risks involved, may serve to increase trust in the exercise itself, although this cannot be guaranteed. While transparency might be cited as a way to increase trust, in the care.data scenario, transparency by way of social media and other forums may well have served to undermine trust by highlighting the polarisation of the opposing 'sides,' illustrating O'Hara's assertion that "Transparency would certainly undermine many trust relationships, by making clear when the trustee's interests were not sufficiently well-aligned with the trustor's, or when she was not deliberating in good faith."[38]

3.3 Example - Care.data and trust

Underestimating concerns about real or perceived privacy risks can be costly, as the UK Government found in relation to the so-called 'care.data' project. The Health & Social Care Information Centre (HSCIC), a new 'body corporate' created by statute,[39] was given powers to require any health or social care body, or persons providing health services or social care, to provide it with any information which it considers 'necessary or expedient' for the purposes of its statutory functions[40] (which include collection and analysis of information as directed by the Secretary of State). The care.data project involved the proposed collection by the Centre of patient information held by General Practitioner (GP) surgeries and the use of that information (in most cases, in a de-identified format) for research, monitoring disease and treatment trends, and for assessing safety and risk. Despite the Centre's statutory power of information acquisition, concerns remained as to whether the acquisition of identifiable patient information was in fact necessary and proportionate for the Centre's functions and therefore whether the processing was fair, and whether, in relation to information originally collected for the purpose of the patient's treatment by the GP (in particular prior to the Centre's creation), onward disclosure to the Centre was compatible with that original

---

[37] See note 3 above at 39.

[38] K. O'Hara, "Transparency, Open Data and Trust in Government: Shaping the Infosphere*"* in *ACM Web Science, Evanston, US, 22-24 June 2012*, at 4.

[39] *Health and Social Care Act* 2012, s. 252.

[40] Ibid, at s. 259.

purpose.[41] In the event, patients were offered an opportunity (communicated via patient information leaflet) to opt-out of the project (of their identifiable information being transferred from their GP to the Centre, of their identifiable information being transferred from the Centre to a third party, or both) provided that they contacted their GP.

The opt-out option and the project's apparently worthy aims did not, however, prevent a campaign of protest over the privacy aspects of the proposed data upload. These were based on the following main concerns:

- First, the perceived inadequacies of the awareness raising activities, with a YouGov poll finding that 67% of respondents had not received the information leaflet;[42]
- Secondly, the re-purposing of patient data and concerns that the planned use of that data by the Centre was incompatible with the original purpose of the data collection by the GPs;
- Thirdly, the under-acknowledgment in the leaflet of the risks associated with the pseudonymisation technique used to de-identify patient data before analysis and release to an applicant organisation. As the NHS itself stated, "in theory, a determined analyst could attempt to re-identify individuals within amber [pseudonymised] data by linking them to other data sets."[43] Anonymisation did not allay concerns about transfer of the data to the Centre in the first place, as the Centre was to acquire identifiable information;
- Fourthly, reported links between the project and use of medical data by insurers, drug companies and others in the commercial sector.[44] This reflected an underlying concern that secondary use of the data would not be limited to matters of immediate benefit to NHS. Solove has said that "Secondary use resembles breach of confidentiality, in that there is a betrayal of the person's expectations when giving out

---

[41] Applicable provisions of the Data Protection Act 1998 include: Schedule 3 para 7(1)(b) 'The processing is necessary for the exercise of any functions conferred on any person by or under an enactment'; Schedule 1 Part 1 para 2 'Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.' For a fuller discussion of the data protection debate around care.data, see for instance S. Pritchett, *Care.data – why should we care?* P. & D.P. 2014, 14(4), 14-16.

[42] P. Bradshaw, "Care.data: trust is on the line", *The Guardian* (11 March 2014) available at http://www.theguardian.com/healthcare-network/2014/mar/11/caredata-nhs-trust-doctor-patient-leaflet (last accessed 26 March 2014).

[43] G. Lewis, "Better information means better care*"* (15 January 2014) available at http://www.england.nhs.uk/2014/01/15/geraint-lewis/ (last accessed 26 March 2014).

[44] S. Borland , A. Dolan. and M. Ledwith, "GPs revolt on patient records: Growing anger at NHS plan to harvest private data", *The Daily Mail* (5 February 2014) available at http://www.dailymail.co.uk/news/article-2552651/GPs-revolt-patient-records-Growing-anger-NHS-plan-harvest-private-data.html (last accessed 26 March 2014).

information. One argument to the contrary is that people should simply expect that their data might be used in different ways when they relinquish it" although this creates an "asymmetry of knowledge problem"[45] a factor that was apparent in the care.data debate;

- Fifthly, fears about technical vulnerabilities and hacking, coupled with the above-mentioned re-identification fears.[46]

Goldacre commented that "it is hard to give the team behind care.data our blind faith: they have been caught red-handed giving false reassurance on the very real – albeit modest – privacy threats posed by the system."[47] From these and similar comments it can be extrapolated that underlying the above concerns was the issue of trust. Although "no panacea for the world's ills,"[48] there is little doubt that trust is an essential ingredient of an individual's attitude towards disclosure of their personal data; "Once trust in data custodians has been eroded, it will be difficult to regain."[49]

An informal analysis of a selection of tweets sent by some proponents and critics of care.data indicates tendencies towards group polarisation,[50] and illustrates both the good and bad aspects of social informational cascades as described by Sunstein, as follows:

> Sometimes cascade effects will eliminate group or public torpor by generating concern about serious though previously ignored problems. But sometimes cascade

---

[45] See above note 31, at 520.

[46] D. Martin, "Hack attack on NHS data 'is inevitable': MP claims relationships could be ended and careers destroyed if medical information is made public", *The Daily Mail* (12 February 2014) available at http://www.dailymail.co.uk/news/article-2557286/Hack-attack-NHS-data-inevitable-MP-claims-relationships-ended-careers-destroyed-medical-information-public.html (last accessed 26 March 2014).

[47] B. Goldacre, *The NHS plan to share our medical data can save lives – but it must be done right* The Guardian, (21 February 2014) available at http://www.theguardian.com/society/2014/feb/21/nhs-plan-share-medical-data-save-lives?CMP=twt_fd (last accessed 26 March 2014).

[48] See above note 12, at 231.

[49] K. El Emam, *Guide to the De-Identification of Personal Health Information* (1st edn, Taylor & Francis Group, Boca Raton, 2013) at 102.

[50] 15 February 2014: tweet from @tkelsey1 "now you are being a bit bonkers: none of us will have access – only NHS commissioners, mostly GPs"; 16 February 2014: tweet from @Jarvis: "@tkelsey 1 has lied consistently about #caredata, despises public opinion and has a vested £ interests"; 16 February 2014: tweet from @0spotz: "It seems likely GCHQ will be the first recipient of #caredata feed"; 17 February 2014: tweet from @owenboswarva "Good to see #caredata defenders are working up their own conspiracy theories (see "Who benefits from mass opt-out")"; 18 February 2014: tweet from @lawrenceberry "Civil libertarians like @medConfidential protest. "That's what they do. That's why they exist. No balance betwn patient lives and protest"; 18 February 2014; tweet from @SigniusNetworks "Clueless selfish people like yourself are the reason i do not want my confidential data in your hands".

> effects will make people far more worried than they should be, or otherwise produce large-scale distortions in private judgments, public policy and the law….The serious risk with social cascades, both informational and reputational, is that they can lead to widespread errors, factual or otherwise.[51]

At time of writing, the upload of data from GP surgeries has been delayed until autumn 2014, limited initially to a trial number of surgeries. This provides an opportunity for the development of what O'Hara has described as "warranted" and "accurately-placed" trust[52] through deliberation. This can be achieved provided that a range of views are represented in such deliberation, that it can be based on sober analysis of facts and the tendency (on both sides) to start from an attitude of suspicion can be overcome. If this does not happen (and doubt must remain as to whether genuine consultation on the care.data project will in fact occur), although a public body may increase its transparency, doubts may remain about the transparency of the transparency process itself. Compliance with the law is not enough: "data protection is not sufficient for preserving privacy, or public trust, or indeed the usability of data."[53]

The care.data scenario illustrates what O'Hara has been described as "the classic type of privacy problem", one where humans find it hard "to balance the tangible benefits and the intangible costs."[54] The potential for damage when sharing personal data is much more difficult to quantify compared to, say, the sharing of physical property. When sharing personal data with an online shopping site or social media service, the benefits to the individual may be tangible but the costs less so. The reverse may be true for sharing personal data with the public sector. The benefits for the individual may not be immediate or obvious, but the risks or *perceived* risks - based on reported data breaches, fear of a 'surveillance' society, concerns that anonymised data could be re-identified, previous experiences of function-creep by Government and what O'Neill has described as 'a culture of suspicion'[55] – are much more apparent. The challenge for the public sector is to convince those whose personal data is being used of the long term or societal benefits of the project (which may not have any immediate impact on, or benefit for, the individual), the consequences of abstaining from contributing to the project (to the individual, everyone else and/or the public good), and the real extent of the risks. How will function creep be prevented and individuals given

---

[51] C.R. Sunstein., *Deliberative Trouble? Why Groups Go to Extremes*, (2000-20011) 10 Yale L.J. 71 at 84.

[52] See above note 38, at 4.

[53] See above note 38 at 8-9.

[54] See above note 12, at 5.

[55] O. O'Neill, 'A Question of Trust' *The BBC Reith Lectures* (2002) available at http://www.bbc.co.uk/programmes/p00ghvd8.

sufficient knowledge and choice over different secondary uses of their data, while allowing public authorities appropriate flexibility in the use of data? As care.data demonstrated, a particularly important risk factor in many data sharing scenarios is the risk of re-identification of anonymised data.

4. Trust, Risk and Anonymisation

Wacks described anonymity as "an important democratic value"[56] as anonymity aids and promotes values such as privacy and free speech. Anonymisation[57] releases the data controller from compliance with the data protection principles of EU law.[58] Anonymisation is important as it enables secondary use of personal data while minimising the privacy risk to individuals. This aspect of anonymisation will be explored further.

A number of studies indicate that the public's attitude towards anonymisation (in the context of data sharing) is relatively positive.[59] The Health Data Exploration Project explored the barriers to using personal health data for research from individuals who track the data about their own health (using wearable devices, mobile apps and social media).[60] Anonymity was 'very' or 'extremely' important to 67% of participants.[61] But anonymisation cannot alleviate all concerns about the security of personal data, because it is not a clear-cut, one-size-fits-all concept. It requires organisations to think about sufficiently anonymising the data, but also to think about how to retain "data utility."[62] Determining whether personal data has been effectively anonymised involves an assessment of risk in order to ensure, as the UK's Information Commissioner advises, that the risk is "remote."[63] For instance using

---

[56] See above note 1, at 24

[57] That is converting personal data into anonymised form in such a way that a living individual can no longer be identified from it (taking into account all the means likely reasonably to be used by anyone receiving the data)

[58] Recital 26 of the *European Data Protection Directive* (95/46/EC) "whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable".

[59] Such as NHS Information Authority in conjunction with The Consumers' Association *Share with care! People's views on consent and confidentiality of patient information* Birmingham, NHS Information Authority (2002); Stone et al "Sharing patient data: competing demands of privacy, trust and research in primary care" (2005) 55(519) *BMJ* 783-789.

[60] Health Data Exploration project, *Personal Data for the Public Good, New Opportunities to Enrich Understanding of Individual and Population Health,* Final Report (March 2014).

[61] See above note 60, 14.

[62] E. Mackey, M. Ellio, "Understanding the Data Environment", (Fall, 2013) 20(1) *XRDS* 37- 39, at 38.

[63] Information Commissioner's Office, *Anonymisation: managing data protection risk code of practice* (November 2012), at 16.

pseudonymised data[64] may create a higher risk of re-identification, although as the ICO pointed out, this does not mean that effective anonymisation through pseudonymisation is impossible.[65] The use of pseudonymisation was one of the main areas of concern in relation to the care.data project, linked with fears around hacking and the transfer of the data to bodies that may have a vested interest in re-identification.[66]

Paul Ohm has argued that "re-identification science exposes the promise made by [privacy/data protection] laws - that anonymization protects privacy - as an empty one."[67] Ohm highlighted "release-and-forget" anonymization, with generalised rather than suppressed identifiers, as of concern, particularly as other "data fingerprints" such as search queries or social media postings can be combined with anonymized data to attempt re-identification.[68] Sweeney et al highlighted the absence of awareness that there is a risk of re-identification as of concern:

> …sharing information about sexual abuse, abortions or depression medication may be liberating for one person yet harmful for another. Further, if the information is shared without the explicit appearance of name or address, a person may be more likely to share the information publically because of the false belief she is anonymous.[69]

Others disagree with Ohm's pessimistic view of re-identification. In her 2013 guidance, Ann Cavoukian restated her opinion that re-identification 'is not an easy or trivial task' and that the most significant privacy risks arise from ineffectively de-identified data. Commenting on Big Data, she said: "As masses of information are linked across multiple sources it becomes

---

[64] See note 63 above, at 51 which states "Pseudonymisation is described by the ICO's Code as 'De-identifying data so that a coded reference or pseudonym is attached to a record to allow the data to be associated with a particular individual without the individual being identified."

[65] See note 63 above, at 21.

[66] The HSCIC website explains that "The information collected from GP records will be linked to information that the organisation holds from hospital records. Once the information has been linked together, the details that identify you, such as your date of birth and postcode, will be removed and replaced by a reference number. This is known as pseudonymisation. This data is also referred to as "potentially identifiable" as there is a small risk that you could be identified. For example, if someone already knew that you had a rare disease and there was only one person in your area who had that disease, then they may realise the information relates to you even though your identifiers are not included." http://www.hscic.gov.uk/article/3918/How-information-collected-from-GP-records-for-caredata-will-be-used-and-shared (last accessed 20 March 2014).

[67] P. Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, (2010) 57 UCLA Law Review 1701 at 1704.

[68] *Ibid* at 1723.

[69] L. Sweeney, A. Abu, J. Winn, *Identifying Participants in the Personal Genome Project by Name* (Harvard College, Cambridge, Massachusetts, 2013) at 1.

more difficult to ensure the anonymity of the information."[70] On the other hand, Big Data could make de-identification easier to achieve: "Smaller datasets are more challenging to de-identify as it is easier to be unique in a small dataset."[71]

Who we believe and how we assess the risk of re-identification in any particular circumstance may well be influenced by who we trust: "…our views of the facts about big risks are often prompted by our politics and behaviour, even as we insist that the rock on which we build our beliefs is scientific and objective, not the least bit personal."[72] A recent study examined how much trust the participants had in information provided by certain categories of people.[73] Information provided by scientists was the most trusted (28% trusted the information "a great deal", 46% "a fair amount", compared to politicians: 1% "a great deal", 7% "a fair amount"). And so it might be expected that scientific anonymisation studies would increase public (and decision-maker's) understanding of the risks of re-identification. It seems that the opposite may sometimes be the case.

Trust cannot fail to be affected by what Daniel Barth-Jones has described as "anxiety-inducing media storms" over recent re-identification research and demonstration attacks, many of which, Barth-Jones argued "particularly because of the way their results have been reported to the public, serve to inherently distort the public's (and, perhaps, policy-maker's?) perceptions of the likelihood of 'real-world' re-identification risks." [74] He also commented on the impact of fear on the ability to assess risk rationally: "when a re-identification attack has been brought to life, like some Frankenstein monster, our assessment of the probability of it actually being implemented in the real-world may subconsciously become 100 percent, which is highly distortive of the true risk/benefit calculus that we face."[75]

---

[70] A. Cavoukian, Information & Privacy Commissioner Ontario, Canada, *Looking Forward – De-identification Developments – New Tools, New Challenges* (May 2013), at 11 available at http://www.privacybydesign.ca/content/uploads/2013/05/de-identifcation-developments.pdf (last accessed 26 March 2014).

[71] *Ibid*, at 12.

[72] M. Blastland and D. Spiegelhalter, *The Norm Chronicles*, Profile Books (London: 2013) at 110.

[73] Ipsos MORI Public Understanding of Statistics, Topline Results, Fieldwork: 9th – 15th April 2013.

[74] D.C. Barth-Jones, *Public Policy Considerations for Recent Re-Identification Demonstration Attacks on Genomic Data Sets: Part 1* (Re-Identification Symposium) available at http://blogs.law.harvard.edu/billofhealth/2013/05/29/public-policy-considerations-for-recent-re-identification-demonstration-attacks-on-genomic-data-sets-part-1-re-identification-symposium/ (last accessed 20 March 2014).

[75] *Ibid*.

Lagos and Polenetsky point out that none of the published attacks have occurred using non-public databases.[76] To maintain acceptable standards, they advocate a two-layer approach – technical de-identification combined with administrative safeguards (such as internal administrative and physical controls, and contractual and legal protections) – arguing that this significantly minimises potential privacy harms. In the care.data debate, the existence of administrative safeguards tended to be dismissed by those critical of the project, commonly on the basis of past data breaches in the health sector.[77]

It should be of concern that the debate around re-identification of anonymised personal data tends to sway between *it's so easy that my toddler could do it*, to *trust us, there's nothing to worry about* with the reality being somewhere in between and context-dependent. This may be a manifestation both of the conceptual quality of privacy (resulting in difficulty articulating the privacy harms)[78] and of the risk-based nature of anonymisation: "The underlying issue is the risk-utility trade-off, and its assessment is inherently imprecise, with a host of difficult to quantify factors. This can be daunting for an organisation that is embarking on a data sharing process."[79] The temptation may be to downplay risks and simplify technical explanations in an attempt to reassure whereas 'people need full information and guidance for action, rather than just reassurance, and their concerns must be taken seriously.'[80]

5. Empirical study: Attitudes to sharing personal data with the public sector – an exploratory research project

Empirical work has been described as "an essential component of governments' efforts to ensure and retain public trust in transparency."[81] With this in mind, the aim of this study was to explore attitudes to sharing personal data with the public sector, in particular, attitudes towards local councils, central government and the NHS and focussing on locational data, minor personal ailments and detailed medical history.

---

[76] Y. Lagos and J. Polonetsky, "Public vs. Nonpublic Data, The Benefits of Administrative Control", (2013) 66 *Stanford Law Review Online* 103.

[77] See note 46 above.

[78] See note 7 above at 480.

[79] D. Smith, P. Singleton, M. Elliot, D. Kalra, *The risks involved in care.data's anonymisation system*, StatsLife (24 March 2014) available at http://www.statslife.org.uk/opinion/1296-the-risks-involved-in-care-data-s-anonymisation-system (last accessed 26 March 2014).

[80] See note 72 above, at 119.

[81] See note 38 above at 1.

5.1 Participants

A sample of 131 adults took part in the study (between May and July 2013). 73% were female (N = 95) and 2% did not state their gender (N = 3). The frequencies and percentages of each age group are shown in Table 1. Medical professionals accounted for 17% of responses (N = 22) with their responses not being significantly different from those of others. Participants were recruited via self-selected sampling, with questionnaires being available in a NHS hospital reception and online.

Table 1
Frequency and Percentage of Age

	Frequency	Percentage
18-24	21	16%
25-34	29	22.1%
35-44	21	16%
45-54	28	21.4%
55-75	28	21.4%
Not Defined	4	3.1%
Total	131	100%

5.2 Materials

A 102 item questionnaire was compiled, both in paper and electronic formats. Questions focused on participant's opinions on: providing location and medical information to local councils, the central government and the NHS; how personal data was used by public organisations; anonymised data and privacy terms and conditions (not discussed here), and demographic information.

5.3 Procedure

Following receipt of NHS Ethics and R&D approval, printed questionnaires made available within Hampshire Hospitals NHS Foundation Trust's outpatients department and the online version was also advertised on the University of Winchester's internal university website and via social media. Participants completed the questionnaire without the aid of researchers. Electronic copies were submitted online upon completion and physical copies were given to NHS staff to be returned to the researchers.

5.4 Limitations

Certain aspects of the current study, which limit how the findings can be applied to the wider population, must be addressed. As this was an exploratory study, with 131 participants, the results cannot be generalised to a wider population. The participants were self-selecting and therefore it is possible that there were more participants with a personal interest or knowledge

of privacy issues than would be found in the general population. Nationality was asked for but no more specific geographical information was requested. As such it cannot be known how representative the views are of different regions of the UK. 51% of participants had received medical treatment from a GP or hospital recently; although this appeared to make little difference to attitudes towards sharing of medical data, it may have been the case that recent treatment resulted in a bias either in favour or against the NHS depending on personal experiences.

Previous research has found significant differences between individuals' expressed privacy concerns, and their willingness to share personal data with companies and organisations (known as the "privacy paradox").[82] The privacy paradox has been found to apply to locational data, whereby there was no correlation between stated views on privacy and behaviour.[83] As such, the findings of questions in the current study, which asked how comfortable participants would be with providing their data to public sector organisations and data being used in various ways, may not be informative of how participants would behave in real world situations.

5.5 Results

5.5.1 'Providing information about yourself to these types of public sector organisations'

The current study was interested in investigating how comfortable people were with sharing their data with specified public sector bodies. The response rates for 'very' and 'fairly comfortable' are displayed in Table 2. A series of Wilcoxon Matched Pairs tests were performed to compare how comfortable participants were with sharing different types of data with a local council, the central government and the NHS (with a Bonferroni correction of .006). Participants were found to be significantly more comfortable with providing locational, minor aliment and detailed health history to the NHS than both local council and central government ($p<.001$), and significantly more comfortable providing location information to their local council than the central government ($p<.001$).

---

[82] P.A. Norber, D.R. Horne, & D.A. Horne "*The privacy paradox: Personal information disclosure intentions versus behaviors"* (2007) 41(1) *The Journal of Consumer Affairs* 100-126.

[83] A. M. Zafeiropoulou, D.E. Millard, C. Webberand K. O'Hara, "Unpicking the privacy paradox: can structuration theory help to explain location-based privacy decisions?" in, *ACM Web Science* 2013 (WebSci '13), *Paris, France, 02 - 04 May 2013.*

Table 2
Combined 'Very Comfortable' and 'Fairly Comfortable' response percentages for providing organisations with personal information.

Location Information			Minor Personal Information		Aliment	Detailed Health History		
Local Council	Central Government	NHS	Local Council	Central Government	NHS	Local Council	Central Government	NHS
53%	38%	65%	24%	21%	86%	12%	13%	85%

A series of Wilcoxon Matched Pairs tests revealed that people were significantly more comfortable providing locational and medical information to the NHS, rather than local council or central government. Other research has found high levels of trust in the health sector: *Health Which?*[84] reported that patients were generally comfortable to have their basic record shared within the NHS for treatment purposes, and Eurobarometer found that 83% of those surveyed indicated that they trust health institutions with their data[85]. This is reflected very strongly in the current findings, with over 85% of participants expressing that they are comfortable with providing the NHS with their medical data.

65% of participants said they would be comfortable providing the NHS with their locational data, information that has no immediate or obvious use for medical purposes. It was also found that participants were significantly more comfortable with providing local councils with locational information than central government.

5.5.2 The impact of data breaches

There have been numerous instances where unwarranted disclosures of personal data have occurred in the health sector resulting in enforcement action by the Information Commissioner's Office (ICO).[86] The ICO has reported that 137 data breach incidents occurred in the health sector in the second quarter of 2013, the highest of any sector (though as the ICO itself points out though, NHS organisations are required to self-report potential

[84] Health Which? NHS National Programme for Information Technology *The public view on electronic health records* (7 October 2003), at 14 available at http://www.providersedge.com/ehdocs/ehr_articles/The_Public_View_on_EHR.pdf (last accessed 26 March 2014).

[85] Special Eurobaromater Report 359, *Attitudes on Data Protection and Electronic Identity in the European Union* (2011) available at http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf (last accessed 20 March 2014).

[86] For instance, the monetary penalty notice of £200,000 issued to NHS Surrey, available at http://www.bbc.co.uk/news/technology-23286231 (12 July 2013) (last accessed 20 March 2014).

data breaches).[87] Only 8% of participants in the current study expressed willingness to share personal data after a security breach, one participant commenting that "*I am more concerned over data security (i.e., breaches / loss of data by central and local government) than I am about deliberate sharing*." Yet they were most comfortable with providing the NHS with personal data despite this sector's records on data breaches. These findings beg the questions – are the public aware of data breaches in the public sector? Or do people behave differently to their expressed views?

5.5.3 Comparison to websites and search engines

The comparison of the current results against previous research conducted by IPSOS-Mori on behalf of Osborne Clarke[88] suggest that people are more comfortable providing websites and search engines with their medical information than the local or central government (see Table 3). When people provide websites or search engines with medical data, it is most likely that this is to investigate and locate information about an ailment, thus providing the individual with an immediate benefit perhaps not so apparent when providing data to central or local government. Is it the immediate benefit that results in these comfort levels or perhaps the level of control over the act of sharing?

Table 3
Combined 'Very Comfortable' and 'Fairly Comfortable' response percentages for providing organisations with personal information. Search Engines and Website data taken from Osborne Clarke's study 'The data gold rush: Growing and protecting your position in the data ecosystem'

Minor Personal Aliment Information			Detailed Health History		
Local Council	Central Government	Search Engines & Websites	Local Council	Central Government	Search Engines & Websites
24%	21%	52%	12%	13%	22%

5.5.3 Comfort with public sector organisations collecting, storing and using personal data - anonymisation

The response rates for 'very' and 'fairly comfortable' are displayed in Table 4. A series of Wilcoxon Matched Pairs tests were performed to compare how comfortable participants

[87] Information Commissioner's Office, Data breach incidents for period 1 July - 30 September 2013, available athttp://www.ico.org.uk/enforcement/trends (last accessed 20 March 2014).

[88] O. Clarke *'The data gold rush: Growing and protecting your position in the data ecosystem* (2012) available at http://www.osborneclarke.com/media/filer_public/4d/7f/4d7f52cb-d504-480d-a067-48e27a4b6be0/oc_digital_business_data_gold_rush_interactive_pdf_2013.pdf (last accessed 20 March 2014).

would be with their data being collected, stored and used by local councils, central government and the NHS when anonymised and when identifiable (with a Bonferroni correction of 0.006). All comparisons were highly significant (p<.001) with an increase in very and fairly comfortable responses, with the exceptions of NHS minor aliment (p=.72) and major health history (p=.32) comparisons (see table 4).

Participants were significantly more comfortable with their data being collected, stored and used by local and central government when data was anonymised than if it was not. Participants were, however, only significantly more comfortable with the NHS collecting, storing and using their locational information if it was anonymised. No significant difference was found for minor personal aliment and for detailed health history in regards to the NHS. This, coupled with the finding that around a quarter of participants did not answer 'comfortable' to the NHS collecting, storing and using personal medical information suggests that some other factor makes some people uncomfortable with the NHS storing medical data about themselves.

Table 4
Combined 'Very Comfortable' and 'Fairly Comfortable' response percentages for different public bodies collecting, storing and using personal information – anonymised and identifiable.

	Location Information			Minor Personal Aliment Information			Detailed Health History		
	Local Council*	Central Government*	NHS*	Local Council*	Central Government*	NHS	Local Council*	Central Government*	NHS
Identifiable	22%	27%	48%	19%	17%	74%	8%	11%	79%
Anonymised	57%	46%	66%	48%	48%	75%	44%	45%	75%

** $p < .001$*

5.5.4 Comfort with public sector organisations sharing personal data - anonymisation

The response rates for 'very' and 'fairly comfortable' are displayed in Table 5. A series of Wilcoxon Matched Pairs tests were performed to compare how comfortable participants would be with their data being shared by local councils, central government and the NHS when anonymised and when identifiable (with a Bonferroni correction of 0.006). All comparisons were highly significant (p<.001) with an increase in very and fairly comfortable responses.

Table 5
Combined 'Very Comfortable' and 'Fairly Comfortable' response percentages for different public bodies sharing personal information - anonymisation.

	Location Information			Minor Personal Aliment Information			Detailed Health History		
	Local	Central	NHS*	Local	Central	NHS*	Local	Central	NHS*

	Council*	Government*		Council*	Government*		Council*	Government*	
Identifiable	7%	6%	20%	3%	3%	22%	2%	2%	22%
Anonymised	38%	38%	47%	35%	34%	54%	28%	31%	50%

** $p < .001$*

Results again show higher levels of comfort with the local and central government sharing medical and locational data when anonymised. But comfort levels for anonymised data were still low (<40%). Participants were also significantly more comfortable with the NHS sharing personal data for location and medical data when it is anonymised. One participant said "*Most data to assist public services can be anonymised but I fear is not*." Another expressed the view that *"If my personal data was anonymised, I would answer 'very comfortable' [to different uses of data] as I believe that public sector orgs should only base their changes or improvements on factual data rather than 'finger in the air' responses"* with another commenting that "*Anonymised data should give enough information if it is collected correctly and still respoct (sic) an individual's right to privacy.*"

The above findings complement research carried out for the Scottish Government by IPSOS-Mori and the University of Edinburgh which found a significant level of concern amongst participants about the potential for hacking of, or unauthorised access to, personal data. [89] A 'significant minority' of participants were sceptical about anonymisation, with their views "underpinned by consideration of high profile data losses and breaches, but also by the perceived ease with which commercial organisations in particular appear to come into possession of individuals' details for use in direct marketing campaigns."[90]

Overall, participants in the current study were more comfortable with their data being collected, stored and used than shared, in all organisation, data and anonymisation combinations. It is worth observing that the NHS received the most comfortable responses from those surveyed, regardless of the type of information or whether it was anonymised. This suggests that of people surveyed, more have a greater trust in the NHS handling their data than local or central government bodies despite the statistics regarding personal data breaches.

Comparing the response rates between Table 2 (providing organisations with personal information) and Table 4 (public bodies collecting, storing and using personal information) reveals that more people are comfortable with local and central government collecting, storing and using personal data (when anonymised) than providing them with the

[89] S. Davidson., C. McLean, S. Treanor, M. Aitken, S. Cunningham-Burley, G. Laurie, C. Pagliari and N. Sethi,, *Public Acceptability of Data Sharing Between the Public, Private and Third Sector for Research Purposes* (Scottish Government Social Research, 2013) at 62.

[90] *Ibid*.

information. More participants are comfortable providing the NHS with their health information than the NHS collecting, storing and using that information (when anonymised).

For nearly all organisation and type of personal data combinations, however, over half of participants did not state that they were comfortable with their data being shared when anonymised, suggesting that how data is used affects people's comfort with their data being shared.

5.5.4 Comfort with personal data being used in different ways

The survey asked about different purposes for data usage; the results have been grouped below using the following general categories: developing services, sharing, and crime and monitoring.

Table 6
Combined 'Very Comfortable' and 'Fairly Comfortable' response percentages for different public bodies using personal data for developing services

	Local Council	Central Government	NHS
To improve delivery of public services	67%	59%	85%
To help the public sector create a new service that will benefit society as a whole	60%	58%	67%
To provide me with an individual service that will help me in my daily activities	61%	54%	76%
To help the public sector plan for the future	72%	64%	80%
To help the public sector make money for the benefit of public services	22%	23%	27%

Participants were more comfortable with their data being used to improve current services than creating a new service across all three organisations. Local councils received slightly more comfortable responses than central government (table 6).

Again, across all three public organisations, participants were more comfortable with their data being used to improve current services rather than providing themselves with a service that will help them specifically, suggesting an element of altruism or alternatively that people are less comfortable with an individual service due to the identification or profiling implications.

The majority of participants were comfortable with their data being used to improve the delivery of public services and to help the public sector plan for the future, yet only around a quarter were happy with their data being used to help the public sector make money for public services. Osborne Clark's research[91] found that even fewer people were comfortable with private organisations using personal data to generate money (7%).

Participants were also more comfortable with their data being used to improve services on a local council level than a central government level, although this difference was minimal in relation to creating a new service and generating money for the public sector.

Table 7
Combined 'Very Comfortable' and 'Fairly Comfortable' response percentages for different public bodies sharing personal data with other parties

	Local Council	Central Government	NHS
To share with other public sector organisations	27%	28%	32%
To share with researchers	34%	34%	50%
To share with commercial organisations	2%	8%	8%

Sharing of data with other public sector organisations was found to be generally unfavourable across all organisations investigated, with 'comfortable' responses ranging from a quarter to a third of responses. This result supports previous findings from a New Zealand study, which found that far greater numbers of people were comfortable with their data being shared within a health organisation than shared with organisations external to the health organisation.[92]

Of the different sharing information questions, sharing data with researchers received the highest response rate, although it was only 50% for the NHS and lower for local council and central government (see Table 7). El Emam has argued that people are uncomfortable with their information being used for secondary purposes if people do not trust the organisation originally collecting the data.[93] The results of the current study support this proposition, as

[91] See note 88 above.

[92] R. Whiddett,, I. Hunter,J. Engelbrech,, & J. Handy, "Patients' attitudes towards sharing their health information" (2006) 75 *International Journal of Medical Informatics* 530-541.

[93] See note 49 above at 101.

the NHS generally received more comfortable responses than local or central government in this context. It is however worth pointing out again that only half of people surveyed said they would be comfortable with the NHS sharing their data with researchers. El Emam also commented that people "often cite privacy and confidentiality concerns and a lack of trust in researchers as reasons for not wanting their health information used for research."[94] It would be of interest to investigate whether the type of research being conducted would affect the comfort responses to this question. For example, would participants be more comfortable with the NHS sharing medical information with medical researchers rather than demographic researchers? Anonymisation may well also affect comfort levels relating to research. In a 2003 study conducted in the US, it was found that up to 86% of people surveyed were comfortable with their data being used in a health database of anonymised information for research, but only 35% were comfortable if the database contained identifiable data.[95] The recent Health Data Exploration project made similar findings: 78% of respondents answered 'probably would' or 'definitely would' when asked if they would be willing to share personal health data with researchers if anonymised.[96]

Sharing data with commercial organisations was met with few comfortable responses (all less than 10%). This is in line with previous research. A US study by Grimes-Gruczka and Gratzer[97] found that people were overwhelmingly against third parties gaining their health information, with 88% stating that they were not willing to share health data with advertisers or marketers. The Health Data Exploration project found that many participants would be more likely to share their data if they knew that it would be only be used for 'public good research' with 13% of respondents mentioning an aversion to commercial or profit-making use of their data.[98]

---

[94] See note 49 above at 101.

[95] N. Kass, M. Natowicz, S. Hull, et al. "The use of medical records in research: what do patients want?" (2003) 31 *Journal of Law, Medicine and Ethics* 429-433.

[96] See note 60 above at 13.

[97] T. Grimes-Gruczka, C. Gratzer, *The Institute for the Future Ethics Survey of Consumer Attitudes about Health Web Sites*, California Health Care Foundation (2000).

[98] See note 60 above at 13.

Table 8
Combined 'Very Comfortable' and 'Fairly Comfortable' response percentages for preventing/investigating crime, fraud/tax evasion and monitoring

	Local Council	Central Government	NHS
To help the public sector prevent or investigate crime	70%	66%	63%
To help the public sector prevent or investigate fraud and tax evasion	65%	60%	55%
To monitor an individual's use of all public services	21%	25%	31%

The majority of participants were willing for their data to be used to help prevent or investigate crime, fraud and tax evasion. In relation to these questions, the NHS received fewer comfortable responses than with the local or central government, perhaps reflecting an inability to comprehend how medical data or the NHS itself might be connected with such activities. Furthermore it is interesting to note the small yet uniform decrease in comfortable responses between the two questions, with *help to investigate fraud and tax evasion* receiving 5-8% fewer comfortable responses. Perhaps fraud and tax evasion are regarded as less serious crimes, with some people more willing to make the trade-off between privacy and assisting in criminal investigation with regard to more serious crimes.

Between a fifth and a third of participants said they would be comfortable with their data being used to monitor their use of all public services, the lowest response in this section, hardly surprising given the overtones of 'Big Brother' evident nowadays in the word 'monitor.' It would be interesting to explore whether use of a more neutral word would have changed the results.

Lastly, some participants expressed the view that whether they would be happy with their information being shared was dependent on a number of factors including: knowing what information was required, how it was being collected, whom it was being shared with, whether data was anonymised and whether they were asked to give their consent.

6. Conclusions

Returning to the article's initial question – can it be concluded that the necessity of public services results in readiness in individuals to share personal data and sacrifice a certain level of privacy? – the short answer must be *no*. Both the literature review and the current exploratory study suggest that the benefits-versus-costs problem is particularly significant:

the more tangible and/or immediate the benefit, the stronger the correlation to (and possibly the cause of) comfort in data sharing.

The inevitable follow-on question is how can the public sector allow people to see and understand the benefits of personal data sharing (and why, in some circumstances less privacy for some might be justifiable) in the same way that people appear able to do for themselves in relation to commercial transactions? This becomes a particular challenge when the data sharing involves misunderstood and mistrusted elements such as interaction with the commercial sector or the use of bulk datasets for the generation of intelligence. People surveyed in the current study showed a high level of willingness to provide their information to improve public services but were more reluctant to do so when money was mentioned, or when data sharing was involved, even with other public bodies.

In projects such as care.data, in which individual control over personal data is proposed to be diluted to an opt-out, it should perhaps not be left up to the individual to attempt to balance hard-to-define worries about privacy with an often equally hard-to-define societal aim; rather the public sector should ensure what Nissenbaum has called a "context-appropriate flow of information"[99] and take steps to minimise any connected risks – important ones being function creep, incompatibility with the original data collection purpose and re-identification – and be better at communicating the safeguards.

The current exploratory study showed that comfort levels with all public sector organisations dropped when participants were asked about organisations *sharing* personal data, with anonymisation making less of a difference to comfort levels than was expected. More could be done by some of those commenting on anonymisation neither to downplay nor to overplay the re-identification risks, despite being less headline-grabbing. Differences of opinion regarding the risk of re-identification in any given case are inevitable but as a recent letter to the Times expressed, "The only solution is transparency. We need to know who has access to our data, in what form and what they're doing with it."[100] The law has a role to play and a valuable avenue of future investigation would be a comparison of UK law in this area with that of the US, in particular the 'expert determination' and 'safe harbor' methods of de-identification under the Health Insurance Portability and Accountability Act 1996[101]; would

---

[99] See note 16 above at 187.

[100] R. Anderson., I. Brown, J. Crowcroft, F. Fisher, D. Korff , *Safeguarding patient data* (20 March 2014) available at http://www.timeshighereducation.co.uk/comment/letters/safeguarding-patient-data/2012114.article (last accessed 27 March 2014).

[101] Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, OCR (November 26, 2012) available at http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf (last accessed 27 March 2014).

the safe harbor method, under which specified identifiers are removed, increase trust in anonymisation and what would be the impact on data utility?

A final thought about localism. The exploratory study consistently demonstrated that the more tangible the public service, the more trust is shown in it. While the NHS is a national organisation, people interact with it face-to-face and therefore it is perhaps the most tangible of the public services dealt with in the research and the one where the benefit of data sharing may be the most apparent. The pattern of trust held even for location data which is not of immediate apparent use for medical purposes. Research has recognised that trust in government is highest at the local level,[102] a factor evident from the results of the current exploratory study. The trust displayed in the NHS may result from its classification by individuals as a local body (it is possible to put a face to a name) and perhaps also from trust in its 'institutional expertise' and image.[103] By contrast, in the care.data project, the HSCIC is an unknown quantity and although part of the NHS, it may be regarded as a 'megastructure'[104] undefinable, impersonal and therefore a little scary. Will the current trend within the NHS for amalgamation and rationalisation, and the creation of larger and often less local bodies based on data-driven economies of scale have a detrimental effect on public trust in the NHS? If trust in, and acceptance of, data sharing initiatives might be improved by a bottom up, local institution-led approach, then decision-makers may be faced with a challenge to balance public confidence and efficient medical care. Future work may wish to investigate how this balance could be achieved.

---

[102] C.A. Cooper,, H. Gibbs Knotts. K.H. Brennan, "The Importance of Trust in Government for Public Administration: The Case of Zoning" (May/June 2008) *Public Administration Review* 459-468, at 459.

[103] C.J. Tolbert, K. Mossberger, "*The Effects of E-Government on Trust and Confidence in Government"* (May/June 2006) *Public Administration Review*354-369, at 356.

[104] R.K. Vischer,*Subsidiarity as a Principle of Governance: Beyond Devolution* (2001) 35 *Indiana Law Review* 103-142, at 116.